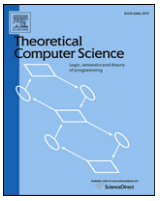
Theoretical Computer Science 405 (2008) 75–87

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs



Topology in information theory in topology

Keye Martin

Center for High Assurance Computer Systems (Code 5540), Naval Research Laboratory, Washington DC 20375, United States

ARTICLE INFO

Keywords:
Domain theory
Quantum information
Topology
Information theory
Timing channels

ABSTRACT

We prove that timed capacity in information theory is a Euclidean continuous function of noise. This is a result based on topological methods that benefits work in information theory. Then we show that binary timing capacity is a measure of distance which yields the Euclidean topology on the unit interval, despite the fact that it does not satisfy the triangle inequality. This is a result based on information theoretic methods that benefits topology. These results have important applications in an area known as information hiding, in the study of quantum communication and in domain theory. They appear to raise fundamental questions about the nature of distance itself.

1. Introduction

In the last year, the author has been pursuing ideas intended to liberate information theory as much as possible from the equations and inequalities which make it a very difficult subject to apply in practice. Algebraic, geometric and order theoretic ideas have been discovered which make it much easier to study the process of communication. This paper is about the use of another subject not traditionally used in information theory, topology, and some of the problems that it has played a key role in solving. Conversely, we will also see that information theoretic ideas may hold a place in topology.

Topological ideas have proven themselves important in information theory lately: in reducing the threat of covert communication [6], in studying the limiting behavior of the capacity achieving distribution [5], in comparing channel behavior [6]. In one manner or another, all of these applications require knowing that the amount of information one can get through a channel varies continuously as a function of noise. Thus far, however, this continuity has only been verified in the case of binary channels [6]. In this paper, we prove that the capacity of a channel with any number of inputs and any number of outputs is continuous, and in fact, we show that this also holds for timing channels with arbitrary inputs and outputs. The result provides a rigorous mathematical foundation for results in [5] and is also used to do new things, such as characterize when the equation arising in the capacity reduction problem is solvable for general timing channels. So we see topology as a useful tool in information theory.

Unexpectedly, though, we have also recently discovered that ideas from information theory have relevance in topology: the most important topology in mathematics, the Euclidean topology, is determined by *channel capacity*. This follows from the more fundamental fact that capacity is a Lebesgue measurement on the interval domain, a result whose proof reveals a profound connection between the study of measurement in domain theory, and the remarkable result of Majani and Rumsey in information theory [2]. The fact that capacity is a measurement makes it possible to reason about capacity by only examining probabilities in the noise matrix of a channel. We illustrate the power of this idea by studying the effect that amplitude damping has on communication with qubits: connecting quantum information's intuitive use of the word 'noise' to information theory's more rigorous use of it.

E-mail address: keye.martin@nrl.navy.mil.

0304-3975/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.06.027

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2008		**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Topology in information theory in topology	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Center for High Assurance Computer Systems (Code 5540),Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We prove that timed capacity in information theory is a Euclidean continuous function of noise. This is a result based on topological methods that benefits work in information theory. Then we show that binary timing capacity is a measure of distance which yields the Euclidean topology on the unit interval, despite the fact that it does not satisfy the triangle inequality. This is a result based on information theoretic methods that benefits topology. These results have important applications in an area known as information hiding, in the study of quantum communication and in domain theory. They appear to raise fundamental questions about the nature of distance itself.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**13**	

2. Topology in information theory

2.1. The Euclidean continuity of capacity

In this section, we establish a continuity principle for information theory that explains why the largest amount of information one can transmit through a communication channel varies continuously as a function of noise. We will formulate the principle for functions of the form $f : X \times Y \to Z$ where Z is a bicontinuous poset and X and Y are spaces in anticipation of future work in relativity [7]. Let us now review the definition of a bicontinuous poset.

Definition 2.1. Let $(P, \sqsubseteq)$ be a partially ordered set. A nonempty subset $S \subseteq P$ is *directed* if $(\forall x, y \in S)(\exists z \in S)\ x, y \sqsubseteq z$. The *supremum* of $S \subseteq P$ is the least of all its upper bounds provided it exists. This is written $\bigsqcup S$.

Dually, a nonempty $S \subseteq P$ is *filtered* if $(\forall x, y \in S)(\exists z \in S)\ z \sqsubseteq x, y$. The *infimum* $\bigwedge S$ of $S \subseteq P$ is the greatest of all its lower bounds provided it exists.

Definition 2.2. For a subset X of a poset P, set

$$\uparrow X := \{y \in P : (\exists x \in X)\, x \sqsubseteq y\} \ \&\ \downarrow X := \{y \in P : (\exists x \in X)\, y \sqsubseteq x\}.$$

We write $\uparrow x = \uparrow \{x\}$ and $\downarrow x = \downarrow \{x\}$ for elements $x \in X$.

Definition 2.3. For elements x, y of a poset, write $x \ll y$ iff for all directed sets S with a supremum,

$$y \sqsubseteq \bigsqcup S \Rightarrow (\exists s \in S)\, x \sqsubseteq s.$$

We set $\Downarrow x = \{a \in P : a \ll x\}$ and $\Uparrow x = \{a \in P : x \ll a\}$.

For the symbol "$\ll$" read "approximates".

Definition 2.4. A *basis* for a poset P is a subset B such that $B \cap \Downarrow x$ contains a directed set with supremum x for all $x \in P$. A poset is *continuous* if *it has a basis. A poset is* ω-continuous *if it has a countable basis.*

A possible objection to our notion of 'approximation' is that it is biased toward suprema; one could similarly define a form of approximation in terms of infima. All notions of bias are removed when we require that these notions coincide.

Definition 2.5. A continuous poset P is *bicontinuous* if

- For all $x, y \in P, x \ll y$ iff for all filtered $S \subseteq P$ with an infimum,

$$\bigwedge S \sqsubseteq x \Rightarrow (\exists s \in S)\, s \sqsubseteq y,$$

  and
- For each $x \in P$, the set $\Uparrow x$ is filtered with infimum x.

Examples of bicontinuous posets are $\mathbb{R}$ and $\mathbb{Q}$ with the usual order.

Definition 2.6. On a bicontinuous poset P, sets of the form

$$(a, b) := \{x \in P : a \ll x \ll b\}$$

form a basis for a topology called *the interval topology*.

That the open intervals (a, b) form a basis for a topology follows from bicontinuity and the fact that continuous posets are *interpolative*: if $x \ll y$ in a continuous poset P, then there is $z \in P$ with $x \ll z \ll y$. On a bicontinuous poset P, the interval topology is Hausdorff, and $\leq$ is a closed subset of P^2. The sets $\uparrow x$ and $\downarrow x$ are closed while the sets $\Uparrow x$ and $\Downarrow x$ are open.

Lemma 2.7. *If $f : X \to Y$ is a continuous function from a compact space to a bicontinuous poset and the image of f is directed, then*

$$f(w) = \bigsqcup_{x \in X} f(x)$$

for some $w \in X$.

Proof. First, we need to show that the indicated supremum exists. The set $K = f(X)$ is a compact subset of Y by the continuity of f. By assumption, K is directed. For each $t \in K$, the set

$$K_t := \uparrow t \cap K$$

is nonempty because it contains t and compact because it is a closed subset of K, using that $\uparrow t$ is a closed subset of Y. By the directedness of K, the collection $\{K_t : t \in K\}$ is a filtered intersection of nonempty compact subsets of K. By the compactness of K,

$$\bigcap_{t \in K} K_t \neq \emptyset.$$

Let u be a point in this intersection. Then $t \leq u$ for all $t \in K$ and $u \in K$. Thus, $u = \bigsqcup K$. Because $u \in K$, there is $w \in X$ with $f(w) = u$. □

We should point out that a compact set in a bicontinuous poset need not be directed. For instance, take the disjoint union $Y = (0, 1/2) \cup (1/2, 1)$ ordered so that elements compare in the usual order only when they belong to the same set. Then $\{1/4, 3/4\}$ is a compact subset with no upper bound. Notice that Y is even globally hyperbolic.[1]

By taking $Y = \mathbb{R}$, we obtain the well known fact from calculus that a continuous real valued function on a compact set assumes a maximum value.

Definition 2.8. If $f : X \times Y \to Z$ is a function that maps into a poset Z, we say that its *image is directed in its second variable* if $f(\{x\} \times Y)$ is a directed subset of Z for each $x \in X$.

Theorem 2.9. *Let X and Y be spaces with Y compact, and let Z denote a bicontinuous poset. If $f : X \times Y \to Z$ is a continuous function whose image is directed in its second variable, then $f^* : X \to Z$ given by*

$$f^*(x) = \bigsqcup_{y \in Y} f(x, y)$$

is a continuous function.

Proof. For a fixed x, the function $f_x : Y \to Z$ given by $f_x(y) = f(x, y)$ is continuous, as the restriction of the continuous f, and has a compact domain since $\{x\} \times Y$ is the product of compact spaces. By Lemma 2.7, f^* is a well-defined function.

Suppose now that $f^*(x) \in (a, b)$. Because $\bigsqcup f(\{x\} \times Y)$ actually belongs to $f(\{x\} \times Y)$, as shown in the previous argument, we know that for some $v \in Y$, $f^*(x) = f(x, v)$. By the continuity of f, there are open sets $U \subseteq X$, $V \subseteq Y$ with $(x, v) \in U \times V$ such that

$$f(x, v) \in f(U \times V) \subseteq (a, b)$$

Thus, if we have any $t \in U$, then

$$a \ll f(t, v) \leq \bigsqcup_{y \in Y} f(t, y) = f^*(t)$$

which means $f^*(t) \in \uparrow\!\!\!\uparrow a$ when $t \in U$. Now we want to find an open set $W \subseteq X$ with $x \in W$ and $f^*(t) \in \downarrow\!\!\!\downarrow b$ for $t \in W$. If we do, then we will have $f^*(t) \in \uparrow\!\!\!\uparrow a \cap \downarrow\!\!\!\downarrow b = (a, b)$ for all $t \in U \cap W$, which will establish the continuity of f^* since the sets (a, b) are a basis for the interval topology on Z.

Let $I = \{U \subseteq X : U \text{ is open } \& \, x \in U\}$. By way of contradiction, we can assume that

$$(\forall i \in I)(\exists x_i \in i) f(\{x_i\} \times Y) \cap (Z \setminus \downarrow\!\!\!\downarrow b) \neq \emptyset$$

(Otherwise, there is $i \in I$ with $f(\{t\} \times Y) \subseteq \downarrow\!\!\!\downarrow b$ for all $t \in i$, and since we always know that $f^*(t) \in f(\{t\} \times Y)$, the proof would be finished.) By Lemma 2.7, there is $w_i \in Y$ with $f^*(x_i) = f(x_i, w_i)$. The set I is a directed poset under the order $U \leq V \Leftrightarrow V \subseteq U$. Thus, by the compactness of Y, the net[2] (w_i) has a subnet specified by a directed subset J of I which converges to a point $w \in Y$. Because the net (x_i) converges to x, the subnet (x_j) also converges to x. Then since $f(x_j, w_j) \in Z \setminus \downarrow\!\!\!\downarrow b$ for all j, the continuity of f and the fact that $Z \setminus \downarrow\!\!\!\downarrow b$ is closed give

$$\lim_{j \in J} f(x_j, w_j) = f(x, w) \in Z \setminus \downarrow\!\!\!\downarrow b.$$

This contradicts $f(x, w) \leq f^*(x) \ll b$. Then there is an open set W with $f^*(t) \ll b$ for all $t \in W$. □

This result can be summarized by saying "if f is directed in a compact variable, then f^* is continuous". One particular application of this theorem, the one we'll be concerned with in the present work, is the case when $Z = \mathbb{R}$:

Corollary 2.10. *If $f : X \times Y \to \mathbb{R}$ is a continuous function with Y compact, then*

$$f^*(x) = \sup_{y \in Y} f(x, y)$$

defines a continuous function from X into $\mathbb{R}$.

A nice aspect of bicontinuous posets is that the order dual of Z, the poset Z^*, is also bicontinuous. Thus, the previous results hold if sup and directed are replaced by inf and filtered. Now we come to the importance of this result in information theory.

Let u denote the noise matrix for an (m, n) channel f i.e. a channel with m inputs and n outputs. Then $u = (u_1, \ldots, u_m)$ is a vector of classical states, each $u_i \in \Delta^n := \{x \in [0, 1]^n : \sum x_i = 1\}$ being a row of the matrix u. If the inputs to the channel $f : \Delta^m \to \Delta^n$ are distributed as $x \in \Delta^m$, then the outputs distribute as

$$f(x) := x \cdot u \in \Delta^n.$$

[1] A poset $(X, \leq)$ is *globally hyperbolic* if it is bicontinuous and if the sets $[a, b] := \{x \in X : a \leq x \leq b\}$ are compact in the interval topology.

[2] Those unfamiliar with the basic properties of nets will find them discussed in the Appendix.

Given times $t = (t_1, \ldots, t_n)$, the mutual information is the function $I_t : \Delta^m \to \mathbb{R}$ given by

$$I_t(x) = \frac{H(f(x)) - \sum_{i=1}^m x_i H(u_i)}{t \cdot f(x)}$$

where H is the base two entropy

$$H(x) = -\sum_{i=1}^m x_i \log_2(x_i).$$

The capacity of an (m, n) timing channel with times $t = (t_1, \ldots, t_n)$ is then the function

$$\mathsf{C}_t : \prod_{i=1}^m \Delta^n \to \mathbb{R}$$

given by

$$\mathsf{C}_t(u) = \sup_{x \in \Delta^m} I_t(u, x)$$

where $I_t(u, x)$ explicitly denotes the dependence of I_t on the noise matrix u. This is called the *timed capacity*. If the times satisfy $t_i = 1$ for all i, the channel f is called *untimed*, and in that case, we abbreviate the capacity to $\mathsf{C}(u)$. Untimed channels, which are what people most often study in the literature, amount to saying that the cost of sending every symbol is the same; the motivation for studying general timing channels is that the cost of sending a symbol in practice can easily depend on the symbol sent.

Theorem 2.11. *Timed capacity is a Euclidean continuous function of noise.*

Proof. The function $I_t : \left(\prod_{i=1}^m \Delta^n\right) \times \Delta^m \to \mathbb{R}$ given by

$$I_t(u, x) = \frac{H(f(x)) - \sum_{i=1}^m x_i H(u_i)}{t \cdot f(x)}$$

is continuous and defined on the product of two compact spaces. By Corollary 2.10, the function $\mathsf{C}_t : \prod_{i=1}^m \Delta^n \to \mathbb{R}$ is also continuous. □

2.2. Discontinuity of the capacity achieving distribution

We have taken an abstract approach to establishing the continuity of timed capacity, so abstract in fact, that it may mislead the reader into believing that the result is simple. However, when looked at from an applied perspective, which we now do, it is a nontrivial fact that timed capacity is continuous. It is important to realize that not all essential quantities in information theory are continuous. For instance, the capacity achieving distribution $x(a, b)$ of a $(2, 2)$ timing channel has no continuous extension to the unit square – and there is irony in this last remark: the discontinuity of $x(a, b)$ follows from the continuity of timed capacity! Let us now explain this remarkable fact.

In the case of a $(2, 2)$ timing channel, the noise matrix u is

$$u = \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}$$

where $a = P(0|0)$, $\bar{a} := 1 - a = P(1|0)$, $b = P(0|1)$ and $\bar{b} := 1 - b = P(1|1)$. Thus, we can think of u as really being a pair of probabilities (a, b). In [5], it is shown that the capacity $\mathsf{C}_t(a, b)$ of a $(2, 2)$ timing channel (a, b) with times $t = (t_1, t_2)$ and $a \neq b$ is given by

$$\mathsf{C}_t(a, b) = \frac{1}{\ln(2)} \cdot \left(\frac{H(a)(b-1) + H(b)(1-a)}{(a-b)} - \ln(f(x)) \right) \cdot \frac{1}{t_1}$$

where x is the unique number in $[0, 1]$ that satisfies

$$e^{-K/(a-b)} (f(x))^{t_2} - (1 - f(x))^{t_1} = 0,$$

with $f(x) = (a - b)x + b$, $K = (b\varepsilon - t_2)H(a) + (t_2 - a\varepsilon)H(b)$, $\varepsilon = t_2 - t_1$ and H the base e entropy

$$H(x) = -x \ln(x) - (1 - x) \ln(1 - x).$$

The capacity of a $(2, 2)$ timing channel is zero iff $a = b$. This characterization of timed capacity is needed in practice so that we know how to calculate it. From the applied perspective then, it is nontrivial that timed capacity varies continuously with a and b on the *entire* unit square: the expression for C_t contains a $1/(a - b)$ term, so its behavior as we approach the

diagonal $\{(x, x) : x \in [0, 1]\}$ is a priori uncertain. Let us now consider the capacity achieving distribution x. By solving for x as a function of a and b,

$$x(a, b) = \frac{1}{a - b} \cdot \left(e^{\beta(a,b)} - b\right), \tag{1}$$

where

$$\beta(a, b) := \frac{H(a)(b - 1) + H(b)(1 - a)}{a - b} - \mathrm{C}_t(a, b)\, t_1 \ln 2.$$

By the continuity of timed capacity, the input distribution which achieves capacity varies continuously as a function of a and b on $\{(a, b) \in [0, 1] \times [0, 1] : a \neq b\}$. The authors of [5] made use of the continuity of timed capacity since they both knew the present author had shown but not published Theorem 2.11. Assuming the continuity of capacity, they were then able to prove:

Theorem 2.12. *The function x has no continuous extension to the entire unit square.*

The proof proceeds by showing that neither

$$\lim_{(a,b)\to(0^+,0^+)} x(a, b) \quad \text{nor} \quad \lim_{(a,b)\to(1^-,1^-)} x(a, b)$$

exists. The notation $(a, b) \to (0^+, 0^+)$ means that we are free to approach the origin along any path provided we are within the unit square; the notation $(a, b) \to (1^-, 1^-)$ says the same of $(1, 1)$. They then establish the following four limits:

- $\lim_{a\to 0^+} x(a, 0) = 1/e$
- $\lim_{b\to 0^+} x(0, b) = 1 - 1/e$
- $\lim_{b\to 1^-} x(1, b) = 1 - 1/e$
- $\lim_{a\to 1^-} x(a, 1) = 1/e$

To give the reader a feel for how it can be possible that $1/e$ and $1 - 1/e$ always arise as limits, *independent* of the values taken on by t_1 and t_2, let us calculate the first one: from Eq. (1) we have,

$$x(a, 0) = (1 - a)^{(1-a)/a} \cdot e^{-\mathrm{C}_t(a,0)t_1 \ln 2}.$$

Because C_t is a continuous function of $(a, b) \in [0, 1] \times [0, 1]$, as $a \to 0^+$, $\mathrm{C}_t(a, 0) \to \mathrm{C}_t(0, 0) = 0$. Thus, $x(a, 0)$ is the product of two functions which have limits as $a \to 0^+$, which means that x has a limit as $a \to 0^+$. This limit is

$$\lim_{a\to 0^+} x(a, 0) = \lim_{a\to 0^+} (1 - a)^{(1-a)/a} \cdot \lim_{a\to 0^+} e^{-\mathrm{C}_t(a,0)t_1 \ln 2} = \lim_{a\to 0^+} \frac{(1 - a)^{1/a}}{1 - a} \cdot 1 = \frac{1}{e} \cdot 1 = \frac{1}{e}.$$

And so x cannot be continuously extended to the entire unit square. In the case of capacity, the $1/(a - b)$ term does not prevent a continuous extension to the unit square, but in the case of the actual distribution $x(a, b)$ that achieves capacity, it does. The discontinuity of $x(a, b)$ becomes even more disturbing when looked at from the practical viewpoint.

Suppose that we have a timing channel with noise matrix u and capacity achieving distribution x. If we vary u ever so slightly in the sense of Euclidean distance, thereby obtaining a new channel with matrix u_ε, it would seem reasonable that the original distribution x would be a good approximation to the capacity achieving distribution x_ε for the new channel u_ε. It is exactly the fact that x is not continuously extendible to the unit square which proves that this is not true! For instance, two positive capacity channels can have noise matrices as close as one likes, and yet their respective capacity achieving distributions can be nearly a maximum distance apart [5]. Moreover, the kind of channels we encounter in a neighborhood of $(0, 0)$ occur naturally in practice [8].

Finally, the numbers $1/e$ and $1 - 1/e$ are not coincidental; we will see them again in the discussion on measurement.

2.3. Solvability of the capacity equation for timing channels

In [6], the capacity reduction problem for untimed binary channels was shown to have a canonical solution by exploiting algebraic structure of channels in place of a second equation. In this problem, one is trying to solve the equation $\mathrm{C}(u) = c_p$ for a noise matrix u given a desired capacity c_p. More generally, one can ask when it is that this equation can be solved for any (n, n) timing channel, we now solve this problem. Continuity even offers an approach to obtaining canonical solutions: let us suppose we measure the amount of noise in the channel by

$$H(u) = \sum_{i=1}^{m} H(u_i).$$

Theorem 2.13. *Let u be the noise matrix of an (n, n) timing channel with times $t = (t_1, \ldots, t_n)$ and $\omega > 1$ be the unique positive solution of $\sum_{i=1}^{n} x^{-t_i} = 1$. Then*

(i) *The equation $\mathrm{C}_t(u) = c_p$ has a solution iff $0 \leq c_p \leq \log \omega$.*
(ii) *If $c_p \in [0, \log \omega]$, then there is a noise matrix u which not only satisfies $\mathrm{C}_t(u) = c_p$ but which also minimizes H. There is also a noise matrix u which maximizes H and solves $\mathrm{C}_t(u) = c_p$.*

Proof. Each part of this proof depends crucially on the fact that capacity is continuous.

(i) The mutual information is always nonnegative and bounded from above by the noiseless version of our channel:

$$0 \leq I_t(u, x) \leq \frac{H(f(x))}{t \cdot f(x)}.$$

Then we have

$$0 \leq c(u) \leq \sup_{x \in \Delta^n} \frac{H(f(x))}{t \cdot f(x)} = \sup_{y \in f(\Delta^n)} \frac{H(y)}{t \cdot y} \leq \sup_{x \in \Delta^n} \frac{H(x)}{t \cdot x} = \log \omega$$

where the equality on the right is due to Shannon [9]. Thus, if we have a solution of $\mathsf{C}_t(u) = c_p$, then $c_p \in [0, \log \omega]$. Now suppose that we have any $c_p \in [0, \log \omega]$. Then for a noise matrix $v = (v_1, \ldots, v_1)$ whose rows are identical, and the identity matrix $w = (e_1, \ldots, e_n)$ which corresponds to the noiseless version of our channel, we have

$$\mathsf{C}_t(v) = 0 \leq c_p \leq \log \omega = \mathsf{C}_t(w).$$

Since c is continuous and the space of noise matrices is also connected, there must be u with $\mathsf{C}_t(u) = c_p$.

(ii) By the continuity of C_t, the set of solutions

$$S = \{u : \mathsf{C}_t(u) = c_p\} = \mathsf{C}_t^{-1}(\{c_p\})$$

is a closed subset of the compact space of noise matrices. Thus, S itself is compact. H is continuous on a compact set so it assumes both a maximum and a minimum. □

Matrices can be chosen to maximize or minimize any continuous quantity, not just entropy.

3. Information theory in topology

3.1. Capacity as a measurement on the interval domain

Recalling the definitions from Section 2.1, we now briefly review domains and measurements.

Definition 3.1. A *dcpo* is a poset in which every directed set has a supremum. A *domain* is a continuous dcpo.

The *interval domain* over [0, 1] is the set of compact subintervals of the unit interval

$$\mathbf{I}[0, 1] = \{[a, b] : a, b \in [0, 1] \,\&\, a \leq b\}$$

ordered by reverse inclusion

$$[a, b] \sqsubseteq [c, d] \Leftrightarrow [c, d] \subseteq [a, b].$$

The poset $\mathbf{I}[0, 1]$ is a continuous dcpo where

- For directed $S \subseteq \mathbf{I}[0, 1]$, $\bigsqcup S = \bigcap S$,
- $x \ll y \Leftrightarrow y \subseteq \mathrm{int}(x)$, and

Notice that $\mathrm{int}(x)$ refers to the interior of the interval x in its relative Euclidean topology, so that $\mathrm{int}[a, b] = (a, b)$ for $a > 0$ and $b < 1$, while $\mathrm{int}[0, b] = [0, b)$ for $b < 1$ and $\mathrm{int}[a, 1] = (a, 1]$ for $a > 0$.

Definition 3.2. The *Scott topology* on a continuous dcpo D has as a basis all sets of the form $\uparrow\!\!\uparrow x$ for $x \in D$.

Example 3.3. A basic Scott open set in $\mathbf{I}[0, 1]$ is

$$\uparrow\!\!\uparrow [a, b] = \{x \in \mathbf{I}[0, 1] : x \subseteq \mathrm{int}([a, b])\}.$$

In the lower half of the unit square, such a set forms a right triangle whose hypotenuse lies along the diagonal, but whose other two sides are removed.

A function $f : D \to E$ between domains is *Scott continuous* if the inverse image of a Scott open set in E is Scott open in D. This is equivalent to saying that f is *monotone*,

$$(\forall x, y \in D)\, x \sqsubseteq y \Rightarrow f(x) \sqsubseteq f(y),$$

and that it *preserves directed suprema*:

$$f\left(\bigsqcup S\right) = \bigsqcup f(S),$$

for all directed $S \subseteq D$. In particular, for the domain $[0, \infty)^*$ of nonnegative reals in their opposite order, it can be shown that a function $\mu : \mathbf{I}[0, 1] \to [0, \infty)^*$ is Scott continuous iff

(1) For all $x, y \in \mathbf{I}[0, 1]$, $x \sqsubseteq y \Rightarrow \mu x \geq \mu y$, and

(2) If (x_n) is an increasing sequence in $\mathbf{I}[0,1]$, then

$$\mu\left(\bigsqcup_{n\geq 1} x_n\right) = \lim_{n\to\infty} \mu x_n.$$

This is the case of Scott continuity that we are most interested in presently:

Definition 3.4. A Scott continuous $\mu : D \to [0,\infty)^*$ is said to measure the content *of* $x \in D$ if for all Scott open sets $U \subseteq D$,

$$x \in U \Rightarrow (\exists \varepsilon > 0)\, x \in \mu_\varepsilon(x) \subseteq U$$

where

$$\mu_\varepsilon(x) := \{y \in D : y \sqsubseteq x \,\&\, |\mu x - \mu y| < \varepsilon\}$$

are called the ε-approximations of x.

We often refer to μ as simply 'measuring' $x \in D$ or as measuring $X \subseteq D$ when it measures each element of X. The last definition, as well as the next, easily extend to maps μ that take values in an arbitrary domain E.

Definition 3.5. A *measurement* $\mu : D \to [0,\infty)^*$ is a Scott continuous map that measures the content of $\ker(\mu) := \{x \in D : \mu x = 0\}$.

The order on a domain D defines a clear sense in which one object has 'more information' than another: a *qualitative* view of information content. The definition of measurement attempts to identify those monotone mappings μ which offer a *quantitative* measure of information content in the sense specified by the order. The essential point in the definition of measurement is that μ measure content in a manner that is consistent with the particular view offered by the order. There are plenty of monotone mappings that are not measurements – and while some of them may measure information content in *some other sense*, each sense must first be specified by a different information order. The definition of measurement can then be seen as a minimal test that a function μ must pass if we are to regard it as providing a measure of information content.

An explicit formula for the capacity [5] of an untimed $(2,2)$ channel is

$$\mathrm{C}(a,b) = \log_2\left(2^{\frac{\bar{a}H(b)-\bar{b}H(a)}{a-b}} + 2^{\frac{bH(a)-aH(b)}{a-b}}\right)$$

where $\mathrm{C}(a,a) := 0$, $H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the base two entropy and $\bar{x} := 1-x$. Capacity can be regarded a function on the interval domain $\mathbf{I}[0,1]$ by setting $\mathrm{C}[a,b] := \mathrm{C}(a,b) = \mathrm{C}(b,a)$. In [6], it is shown that capacity $\mathrm{C} : \mathbf{I}[0,1] \to [0,1]^*$ is Scott continuous from the interval domain to the unit interval in its dual order i.e. capacity decreases as we move up in the order on intervals. What was far from clear at the time though was whether or not capacity was a measurement. We will now prove that it is. Our proof turns on a profound connection between the study of measurement in domain theory, and the stunning result of Majani and Rumsey from information theory:

Theorem 3.6 (*Majani & Rumsey*). *The capacity achieving distribution x of an untimed $(2,n)$ channel with positive capacity satisfies $1/e < x < 1 - 1/e$.*

To see what this result says in the case we are interested in, let $\Delta = \frac{H(a)-H(b)}{a-b}$. Then the capacity achieving distribution for an untimed $(2,2)$ channel is given by

$$x(a,b) = \frac{1}{a-b}\left(\frac{1}{1+2^{\Delta}} - b\right). \tag{2}$$

Thus, by the result of Majani and Rumsey, the quantity $x(a,b)$ always lies in the open interval $(1/e, 1-1/e)$ whenever (a,b) has positive capacity i.e. when a and b are not equal. This remarkable fact was first observed by Silverman in 1955, but without proof. It is a beautiful result, and quite mysterious. Its beauty is capable of obscuring the fact that it can also be a useful tool for problem solving. In our case, it is the key observation needed:

Lemma 3.7. *For any binary channel $(a,b) \in [0,1]^2$,*

$$\mathrm{C}(a,b) \geq \frac{(a-b)^2}{e^2\ln(2)} \geq \frac{(a-b)^2}{6}.$$

Proof. First, let (a,b) be a positive capacity channel. Then $a \neq b$. By the statement of the lemma and the symmetry of capacity, we can assume $a > b$. Consider the base two mutual information

$$I(x) = H(f(x)) - xH(a) - (1-x)H(b).$$

Notice that $I'(x) = H'(fx)(a-b) - (H(a)-H(b))$ and $I''(x) = H''(fx)(a-b)^2$, for $x \in (0,1)$. Let $x(a,b)$ denote the unique capacity achieving distribution for (a,b).

(i) For the unique capacity achieving distribution $x(a,b)$, $I'(x(a,b))=0$, which can be verified by direct substitution if one wishes, since we have a formula for it,
(ii) The point $x(a,b)$ is in $(0,1)$: since $I(0)=0$ and $I(1)=0$, $x(a,b)=0$ or $x(a,b)=1$ would imply that $\mathsf{C}(a,b)=I(x(a,b))=0$,
(iii) $I''<0$ on $(0,1)$, which follows from $\ln(2)\cdot H''(t)=-1/t(1-t)<0$, which itself follows from $H'(t)=\log_2((1-t)/t)$.

Then $I'>0$ on $(0,x(a,b))\neq\emptyset$, so let $x\in(0,x(a,b))$ be any point where $I'(x)>0$. By the mean value theorem,

$$(\exists c\in(0,x))I(x)-I(0)=I(x)-0=I(x)=I'(c)(x-0).$$

Because $I''<0$, I' is strictly decreasing, so $I'(c)>I'(x)$ and thus

$$\mathsf{C}(a,b)\geq I(x)>x\cdot I'(x)=x\cdot(H'(fx)(a-b)-(Ha-Hb)).$$

Now notice that

$$H(a)-H(b)=(a-b)H'(f(x(a,b))).$$

Thus, our lower bound on capacity can be rewritten as

$$\mathsf{C}(a,b)>x(a-b)[H'(fx)-H'(f(x(a,b)))]>0.$$

The function $f(x)=(a-b)x+b$ is strictly increasing since $a>b$, so $f(x)<f(x(a,b))$, and we can once again apply the mean value theorem to $H'(fx)-H'(f(x(a,b)))$ on the interval $[f(x),f(x(a,b))]$ to obtain $d\in(f(x),f(x(a,b)))$ such that

$$H'(fx)-H'(f(x(a,b)))=H''(d)(fx-f(x(a,b)))=H''(d)(a-b)(x-x(a,b)).$$

Our lower bound on capacity is now

$$\mathsf{C}(a,b)>x(a-b)^2(x-x(a,b))H''(d)=\frac{x(a-b)^2(x(a,b)-x)}{d(1-d)\ln(2)}>0.$$

The absolute minimum of $1/t(1-t)$ on $(0,1)$ is 4, so

$$\mathsf{C}(a,b)>\frac{4(a-b)^2(x(a,b)-x)x}{\ln(2)}>0.$$

Now we choose the value of x that maximizes $(x(a,b)-x)x$, which is the midpoint $x=x(a,b)/2\in(0,x(a,b))$, giving us a lower bound on capacity of

$$\mathsf{C}(a,b)>\frac{x(a,b)^2(a-b)^2}{\ln(2)}>0.$$

By the result of Majani and Rumsey, we know that $x(a,b)>1/e$, so

$$\mathsf{C}(a,b)>\frac{(a-b)^2}{e^2\ln(2)}>0.$$

Since we trivially have equality when $a=b$, the proof is finished. □

Recall that capacity can be regarded as a map on the interval domain by setting $\mathsf{C}[a,b]=\mathsf{C}(a,b)$.

Theorem 3.8. *The capacity* $\mathsf{C}:\mathbf{I}[0,1]\to[0,1]^*$ *is a measurement.*

Proof. The length function $\mu[a,b]=b-a$ is a measurement, and so is μ^2, since it arises as the composition of μ and the isomorphism $t\mapsto t^2$ on $[0,\infty)$. Any positive multiple of a measurement is another, so $\mu^2/6$ is a measurement. Finally, because C is Scott continuous [6] and satisfies $\mathsf{C}\geq\mu^2/6$ by the last lemma, C is also a measurement. □

In fact, the proof above also shows that capacity is a *Lebesgue measurement*. Recall that Lebesgue measurements are the measurements which extend to the convex powerdomain, capture metrizability on continuous posets and complete metrizability on continuous domains [4]. Interestingly, capacity is a naturally occurring example of a Lebesgue measurement that does not satisfy any of the following domain theoretic variants of the triangle inequality,

- For all x,y with an upperbound, there is $z\sqsubseteq x,y$ with $\mu z\leq\mu x+\mu y$,
- For all x,y with an upperbound, there is $z\sqsubseteq x,y$ with $\mu z\leq 2\cdot\max\{\mu x,\mu y\}$,

as the following example shows:

Example 3.9. Let $x=[0,1/2]$ and $y=[1/2,1]$. Then x and y have a common upper bound, but since

$$\mathsf{C}(x)=\log_2(5/4)\ \&\ \mathsf{C}(y)=\log_2(5/4)$$

there does not exist $z\sqsubseteq x,y$ satisfying either of the triangle inequality variants since for the only possible $z=[0,1]$, we get

$$\mathsf{C}(z)=1\not\leq\mathsf{C}(x)+\mathsf{C}(y)=2\cdot\max\{\mathsf{C}(x),\mathsf{C}(y)\}=\log_2(25/16)<\log_2(32/16)=\log_2(2)=1.$$

This also shows that capacity does not satisfy the triangle inequality as a function from $[0,1]^2$ to $[0,1]$.

Because capacity violates the triangle inequality, it is not obvious that its ε balls form a basis for a topology; this fact follows from the fact that capacity is a measurement and that measurements always yield a basis for the Scott topology on their kernel. In more detail, given a measurement $\mu : D \to [0, \infty)^*$ on a domain D with a least element, we can define a function $d : D^2 \to [0, \infty)^*$ as

$$d(x, y) = \inf\{\mu z : z \sqsubseteq x, y\}.$$

By [3], the relative Scott topology on $\ker \mu = \{x \in D : \mu x = 0\}$ has as a basis sets of the form

$$B_\varepsilon(x) = \{y \in \ker \mu : d(x, y) < \varepsilon\}$$

where $x \in \ker \mu$ and $\varepsilon > 0$. If we take $D = \mathbf{I}[0, 1]$ and μ to be capacity C on $\mathbf{I}[0, 1]$, then

$$d([a], [b]) = \mathrm{C}([a] \sqcap [b]) = \mathrm{C}[a, b]$$

assuming $a \leq b$. Because we have a homeomorphism $\ker(\mu) \simeq [0, 1]$ between the relative Scott topology and the Euclidean topology, we have proven:

Corollary 3.10. *The untimed capacity* $\mathrm{C} : [0, 1]^2 \to [0, \infty)$ *satisfies*

(i) $\mathrm{C}(a, b) = \mathrm{C}(b, a)$,
(ii) $\mathrm{C}(a, b) = 0$ *iff* $a = b$,

and the sets $\{y \in [0, 1] : \mathrm{C}(x, y) < \varepsilon\}$ *for* $\varepsilon > 0$ *form a basis for the Euclidean topology on* $[0, 1]$.

We will show later that the result above extends to *timed* capacity. But first, let us illustrate the importance of the order theoretic properties possessed by capacity.

3.2. An example from quantum communication

Let $\mathcal{H}$ be the state space for a two dimensional quantum system. Two parties communicate with each other as follows. First, they agree up front on a fixed basis of $\mathcal{H}$, say $\{|\psi\rangle, |\phi\rangle\}$, which can be expressed in some fixed basis $\{|0\rangle, |1\rangle\}$ as

$$|\psi\rangle = a|0\rangle + b|1\rangle \ \ \& \ \ |\phi\rangle = c|0\rangle + d|1\rangle$$

where the amplitudes a, b, c, d are all complex. The state $|\psi\rangle$ is taken to mean '0', while the state $|\phi\rangle$ is taken to mean '1'. The first party, the sender, attempts to send one of these two qubits $|*\rangle \in \{|\psi\rangle, |\phi\rangle\}$ to the second party, the receiver. The second party receives *some* qubit and performs a measurement in the agreed upon basis. The result of this measurement is one of the qubits $\{|\psi\rangle, |\phi\rangle\}$, which is then interpreted as meaning either a '0' or a '1'.

We say *some qubit* because as $|*\rangle$ travels, it suffers an unwanted interaction with its environment, whose effect on density operators can be described as

$$\varepsilon(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger$$

where the operation elements are given by

$$E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix} \ \ \& \ E_1 = \begin{pmatrix} 0 & \sqrt{\lambda} \\ 0 & 0 \end{pmatrix}.$$

This effect is known as *amplitude damping* and the parameter $\lambda \in [0, 1]$ can be thought of as the probability of losing a photon. Thus, the receiver does not necessarily acquire the qubit $|*\rangle$, but instead receives some degradation of it, describable by the density operator $\varepsilon(|*\rangle\langle*|)$.

The probability that '0' is received when '0' is sent is

$$\alpha = P(0|0) = -2|a|^4 p(\lambda) + |a|^2(\lambda + 2p(\lambda)) + 1 - \lambda$$

while the probability that '0' is received when '1' is sent is

$$\beta = P(0|1) = 2|a|^4 p(\lambda) + |a|^2(\lambda - 2p(\lambda))$$

where $p(\lambda) = -1 + \lambda + \sqrt{1-\lambda} \geq 0$. Thus, each choice of basis defines a classical binary channel (α, β). Notice that the probabilities α and β only depend on $|a|^2$ because $|c|^2 = |a|^2$ and $|b|^2 = |d|^2 = 1 - |a|^2$ by the orthogonality of $|\psi\rangle$ and $|\phi\rangle$, and because the initial expressions for α and β turn out to only depend on modulus squared terms. Because the basis is fixed, $|a|^2 \in [0, 1]$ is a constant and we obtain a function $x : [0, 1] \to \mathbf{I}[0, 1]$ of λ given by

$$x(\lambda) = [\beta(\lambda), \alpha(\lambda)].$$

Let us now establish its domain theoretic nature.

Proposition 3.11. *The trajectory* $x : [0, 1] \to \mathbf{I}[0, 1]$ *is Scott continuous.*

Proof. First we prove that $\alpha(\lambda) \geq \beta(\lambda)$, so that we know x actually maps into the interval domain. Because

$$\alpha(\lambda) - \beta(\lambda) = 4p(\lambda)|a|^2(1-|a|^2) + 1 - \lambda \geq 0$$

with the latter being nonnegative because $|a|^2 \leq 1$ and $\lambda \in [0, 1]$. To prove that x is monotone, we show that $\alpha'(\lambda) \leq 0$ and $\beta'(\lambda) \geq 0$. First notice that $p'(\lambda) \leq 1/2$. This implies that

$$\begin{aligned}\alpha'(\lambda) &= (1-|a|^2)(2|a|^2p'(\lambda) - 1)\\ &\leq 1 \cdot (2p'(\lambda) - 1)\\ &\leq 0\end{aligned}$$

and that

$$\begin{aligned}\beta'(\lambda) &= -2|a|^2p'(\lambda)(1-|a|^2) + |a|^2\\ &\geq -|a|^2(1-|a|^2) + |a|^2\\ &= |a|^4\\ &\geq 0\end{aligned}$$

which shows that x is monotone as a trajectory from $[0, 1]$ in its usual order to $\mathbf{I}[0, 1]$. Because x has continuous measure (with respect to the length measurement), x is Scott continuous. □

One valuable aspect of x being Scott continuous is that we can now make precise the connection between quantum information's intuitive use of the word 'noise' and information theory's precise account of it: the quantity $\mathrm{C}(x(\lambda))$ decreases as λ increases i.e. the amount of information that the two parties can communicate decreases as the probability of losing a photon increases. In the extreme cases,

$$x(0) = [0, 1] \;\&\; x(1) = [|a|^2, |a|^2]$$

yielding respective capacities of 1 and 0. There is a more fundamental idea at work in this example and in many others like it: we have learned about capacity by only examining how the probabilities in the noise matrix change, and this more than justifies the domain theoretic approach. Imagine what would happen if we actually tried to calculate $\mathrm{C}(x(\lambda))$ explicitly: we would have to substitute $\alpha(\lambda) = -2|a|^4p(\lambda) + |a|^2(\lambda + 2p(\lambda)) + 1 - \lambda$ for a and $\beta(\lambda) = 2|a|^4p(\lambda) + |a|^2(\lambda - 2p(\lambda))$ for b into the formula

$$\mathrm{C}(a, b) = \log_2\left(2^{\frac{\bar{a}H(b) - \bar{b}H(a)}{a-b}} + 2^{\frac{bH(a) - aH(b)}{a-b}}\right)$$

and then seek to show that the resulting quantity decreases as λ increases. The reader still unconvinced about the merits of the domain theoretic approach, or who believes that the domain theoretic approach is 'not necessary' is more than welcome to provide an alternative, with one caveat: any alternative approach should yield *new results* the way the domain theoretic approach employed above has.

3.3. Timed capacity as a measure of distance

We will show that binary timing capacity $\mathrm{C}_t(a, b)$ is a distance function that yields the Euclidean topology on $[0, 1]$. First though we derive an equation which characterizes the capacity of a binary timing channel as an implicit function of its noise matrix and the channel times. The equation is an interesting result in its own right, but our primary motivation for deriving it is that it allows for derivation of the symmetry relations satisfied by timed capacity, which are ultimately needed to explain it as providing a measure of distance on the unit interval.

Recall from our discussion earlier that the noise matrix u of a $(2, 2)$ timing channel is

$$u = \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}$$

where $a = P(0|0)$, $\bar{a} = 1 - a = P(1|0)$, $b = P(0|1)$ and $\bar{b} = P(1|1)$, so that u can be represented by a pair of probabilities (a, b). The capacity $\mathrm{C}_t(a, b)$ of a binary timing channel (a, b) with times $t = (t_1, t_2)$ and $a \neq b$ is given by

$$\frac{H(a)(b-1) + H(b)(1-a)}{a-b} - \ln(\Phi(a, b)) = \ln(2) \cdot t_1 \cdot \mathrm{C}_t(a, b)$$

where $\Phi(a, b)$ is the unique solution[3] of

$$e^{-K/(a-b)}x^{t_2} - (1-x)^{t_1} = 0$$

on the interval $[0, 1]$, $K = (b\varepsilon - t_2)H(a) + (t_2 - a\varepsilon)H(b)$, $\varepsilon := t_2 - t_1$ and H is the base e entropy

$$H(x) = -x\ln(x) - (1-x)\ln(1-x).$$

It is remarkable that the dependence on $\Phi(a, b)$ can be eliminated.

[3] $\Phi(a, b)$ is *not* the capacity achieving distribution $x(a, b)$, but is related to it by $\Phi(a, b) = (a-b)x(a, b) + b$.

Theorem 3.12. *The capacity of a* $(2, 2)$ *timing channel is* $c_p > 0$ *iff its noise matrix* $u = (a, b)$ *satisfies* $a \neq b$ *and*

$$(*) \qquad e^{(H(a)(b-1)+H(b)(1-a))/(a-b)} \cdot \frac{1}{2^{t_1 c_p}} + e^{(bH(a)-aH(b))/(a-b)} \cdot \frac{1}{2^{t_2 c_p}} = 1.$$

Its capacity is zero iff $a = b$.

Proof. First notice that K can be written as

$$K = t_2[H(a)(b-1) + H(b)(1-a)] + t_1(aH(b) - bH(a)).$$

Let us abbreviate $\Phi(a, b)$ to Φ. If we multiply the equation for c_p by -1 and then exponentiate both sides we get

$$e^{-(H(a)(b-1)+H(b)(1-a))/(a-b)}\Phi = 2^{-t_1 c_p}$$

so raising both sides to the t_2 gives

$$e^{-t_2(H(a)(b-1)+H(b)(1-a))/(a-b)}\Phi^{t_2} = 2^{-t_1 t_2 c_p}.$$

Using the equation which defines Φ and our above remark about K, we see that

$$e^{-t_2(H(a)(b-1)+H(b)(1-a))/(a-b)}\Phi^{t_2} = (1-\Phi)^{t_1} e^{t_1(aH(b)-bH(a))/(a-b)}.$$

Then we must have

$$(1-\Phi)^{t_1} e^{t_1(aH(b)-bH(a))/(a-b)} = 2^{-t_1 t_2 c_p}$$

which gives

$$(1-\Phi) = 2^{-t_2 c_p} e^{(bH(a)-aH(b))/(a-b)}.$$

However, if we solve the c_p equation for Φ, then we see that

$$\Phi = 2^{-t_1 c_p} e^{(H(a)(b-1)+H(b)(1-a))/(a-b)}.$$

Because $\Phi + (1-\Phi) = 1$, we get

$$e^{(H(a)(b-1)+H(b)(1-a))/(a-b)} \cdot \frac{1}{2^{t_1 c_p}} + e^{(bH(a)-aH(b))/(a-b)} \cdot \frac{1}{2^{t_2 c_p}} = 1.$$

For the converse, assume the noise matrix (a, b) of a channel satisfies (*). Let us denote its capacity by $k_p > 0$. We need to prove that $k_p = c_p$. By the work we just did, we know that

$$e^{(H(a)(b-1)+H(b)(1-a))/(a-b)} \cdot \frac{1}{2^{t_1 k_p}} + e^{(bH(a)-aH(b))/(a-b)} \cdot \frac{1}{2^{t_2 k_p}} = 1.$$

Now notice that for constants $\alpha, \beta, t_1, t_2 > 0$, the function

$$g(x) = \frac{\alpha}{2^{t_1 x}} + \frac{\beta}{2^{t_2 x}}$$

is injective since $x < y \Rightarrow g(x) > g(y)$. Thus, for

$$\alpha := e^{(H(a)(b-1)+H(b)(1-a))/(a-b)} > 0 \quad \text{and} \quad \beta := e^{(bH(a)-aH(b))/(a-b)} > 0$$

we have $g(k_p) = g(c_p) = 1$ and hence $k_p = c_p$. □

Let us comment briefly on what makes this result so surprising. If we look at the c_p equation and the Φ equation, it is clear that we can solve each one of them for Φ, and then equate the resulting expressions to obtain a new equation for (a, b). But that equation will depend on $1 - \Phi$. The last theorem shows that this dependence can actually be eliminated! There are two important cases to emphasize:

Example 3.13. *The binary symmetric channel.* In this case, $a = \bar{b} = 1 - p$ and $b = \bar{a} = p$ with p being the probability of a bit flip. The equation relating capacity c_p to noise is then

$$H(p) = \ln(2^{-t_1 c_p} + 2^{-t_2 c_p}).$$

Notice that in the untimed case, when $t_1 = t_2 = 1$, we obtain the well-known result that the capacity of a binary symmetric channel is $1 - H_2(p)$, where H_2 is the base two entropy.

Here a simulation of the binary symmetric channel might involve inputting an image, flipping p percent of its bits, and then displaying the degraded image.

Example 3.14. *The timed Z channel.* In this case we have $b = 0$ and $d = 1$. Then

$$\frac{H(a)}{a} = \ln\left(\frac{2^{-t_1 c_p}}{1 - 2^{-t_2 c_p}}\right)$$

is the equation relating capacity to noise.

Given a vector $t = (t_1, t_2)$ of times for a $(2, 2)$ timing channel, let us write

$$\mathrm{rev}(t) := (t_2, t_1)$$

for the vector whose times are those of t in reverse order.

Proposition 3.15. *The following symmetry properties hold for binary timing channels:*

(i) $\mathsf{C}_t(a, b) = \mathsf{C}_t(b, a)$
(ii) $\mathsf{C}_t(a, b) = \mathsf{C}_{\mathrm{rev}(t)}(\bar{a}, \bar{b})$

Proof. Let

$$f(a, b) = \frac{bH(a) - aH(b)}{a - b} \;\&\; g(a, b) = \frac{\bar{a}H(b) - \bar{b}H(a)}{a - b}.$$

Then $f(a, b) = f(b, a)$ and $g(a, b) = g(b, a)$. These properties along with the previous characterization of capacity give (i). In addition, we also have $g(a, b) = f(\bar{b}, \bar{a})$ and $g(\bar{b}, \bar{a}) = f(a, b)$, which will then give (ii). □

By these two results, we then see $\mathsf{C}_t(a, b) = \mathsf{C}_{\mathrm{rev}(t)}(\bar{b}, \bar{a})$ and thus, without loss of generality, we can always assume that a timing channel satisfies $a \geq b$ and $t_2 \geq t_1$, if all we are interested in is its capacity. Thus, the restriction to nonnegative binary channels, which we made in [6], can be made in the timed case as well. This may be quite a useful thing to know.

Timed capacity does not satisfy the triangle inequality, for the simple reason that the triangle inequality fails in the untimed case ($t_1 = t_2 = 1$).

Theorem 3.16. *For fixed times $t_1, t_2 > 0$, the timed capacity $\mathsf{C}_t : [0, 1]^2 \to [0, \infty)$ satisfies*

(i) $\mathsf{C}_t(a, b) = \mathsf{C}_t(b, a)$,
(ii) $\mathsf{C}_t(a, b) = 0$ *iff* $a = b$,

and the sets $\{y \in [0, 1] : \mathsf{C}_t(x, y) < \varepsilon\}$ for $\varepsilon > 0$ form a basis for the Euclidean topology on $[0, 1]$.

Proof. By the Euclidean continuity of C_t, all timed capacity open balls are Euclidean open. Given any Euclidean open ball, it contains a timed capacity open ball around any point it contains since

$$\mathsf{C}_t(a, b) \geq \frac{\mathsf{C}(a, b)}{\max\{t_1, t_2\}}$$

and *now* we see that the timed capacity open balls form a basis for a topology and that that topology must be the Euclidean topology on $[0, 1]$. The other properties follow from Proposition 3.15. □

4. Capacity in place of distance?

We consider a few cases where one can replace Euclidean distance by capacity.

4.1. A reformulation of measurement

By simple rescaling, any map $\mu : D \to [0, \infty)$ can be assumed to map into $[0, 1]$. To determine whether a map $\mu : D \to [0, 1]^*$ is a measurement requires consideration of sets of the form

$$\mu_\varepsilon(x) := \{y \in D : y \sqsubseteq x \,\&\, |\mu x - \mu y| < \varepsilon\},$$

However, because capacity also yields the Euclidean topology, we can use the sets

$$\mu_\varepsilon(x) := \{y \in D : y \sqsubseteq x \,\&\, \mathsf{C}(\mu x, \mu y) < \varepsilon\}.$$

We could have put any 'semimetric' in place of Euclidean distance, but we put capacity because it has an intriguing interpretation: objects x and y serve to define probabilities μx and μy which define the noise matrix for a binary channel; the capacity of the resulting channel is a measure of distance between x and y. That is, the amount of information that can be communicated from one point in space to another provides a measure of distance that is capable of yielding the space's topology. This, for instance, is what happens with the unit interval.

4.2. Topology of manifolds

The restriction of capacity from the unit square to $(0, 1)^2$ is a distance function which yields the Euclidean topology on $(0, 1)$. For $x, y \in (0, 1)^n$, the function

$$C(x, y) = \sum_{i=1}^{n} C(x_i, y_i)$$

yields the Euclidean topology on $(0, 1)^n \simeq \mathbb{R}^n$. Thus, a pair of n tuples is understood as a defining an n tuple of binary channels (x_i, y_i), and so the topology of a manifold can be understood as stemming from this function (locally). In particular, this is true of spacetime.

5. Questions

When is binary timed capacity monotone?

For further reading

[1], [10], [11], [12].

Appendix. Topology

Nets are a generalization of sequences. Let X be a space.

Definition A.1. *A* net *is* a function $f : I \to X$ where I is a directed poset.

A subset J of I is *cofinal* if for all $\alpha \in I$, there is $\beta \in J$ with $\alpha \leq \beta$.

Definition A.2. Let $f : I \to X$ be a net with a function $g : J \to I$ such that J is a directed poset and

- For all $x, y \in J, x \leq y \Rightarrow g(x) \leq g(y)$
- $g(J)$ is cofinal in I.

The function $f \circ g : J \to X$ is called a *subnet* of f.

Definition A.3. A net $f : I \to X$ *converges* to $x \in X$ if for all open $U \subseteq X$ with $x \in U$, there is $\alpha \in I$ such that

$$\alpha \leq \beta \Rightarrow f(\beta) \in U$$

for all $\beta \in I$. This is written $f_i \to x$.

The following are all standard results of basic topology:

Theorem A.4. (i) *If (x_i) is a net that converges to x, then so does each subnet of (x_i).*
(ii) *If we have nets $x_i \to x \in X$ and $y_i \to y \in Y$, then the net $z_i = (x_i, y_i) \to (x, y) \in X \times Y$.*
(iii) *A function $f : X \to Y$ is continuous iff for each $x \in X$ and each net (x_i) that converges to x, the net $(f(x_i))$ converges to $f(x)$.*
(iv) *A space X is compact iff every net $f : I \to X$ has a convergent subnet.*

References

[1] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, 1991.
[2] E.E. Majani, H. Rumsey, Two results on binary-input discrete memoryless channels, Proceedings of the IEEE International Symposium on Information Theory (1991) 104.
[3] K. Martin, A triangle inequality for measurement, Applied Categorical Structures 11 (1) (2003).
[4] K. Martin, Fractals and domain theory, in: Mathematical Structures in Computer Science, Vol. 14, issue 6, Cambridge University Press, 2004.
[5] K. Martin, I.S. Moskowitz, Noisy Timing Channels with Binary Inputs and Outputs, in: Lecture Notes in Computer Science, Information Hiding 2006, Springer-Verlag, 2006.
[6] K. Martin, I.S. Moskowitz, G. Allwein, Algebraic information theory for binary channels, Electronic Notes in Theoretical Computer Science 158 (2006) 289–306.
[7] K. Martin, P. Panangaden, A domain of spacetime intervals in general relativity, Communications in Mathematical Physics (2006).
[8] I.S. Moskowitz, An analysis of the timed Z-channel, IEEE Transactions on Information Theory 44 (7) (1998).
[9] Ira S. Moskowitz, Allen R. Miller, Simple Timing Channels. In: IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, CA, May 16–18, 1994, pp. 56–64.
[10] M. Nielsen, I. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2000.
[11] C.E. Shannon, A mathematical theory of communication, Bell Systems Technical Journal 27 (1948) 379–423. 623–656.
[12] R.A. Silverman, On binary channels and their cascades, IEEE Transaction on Information Theory (1955) 19–27.